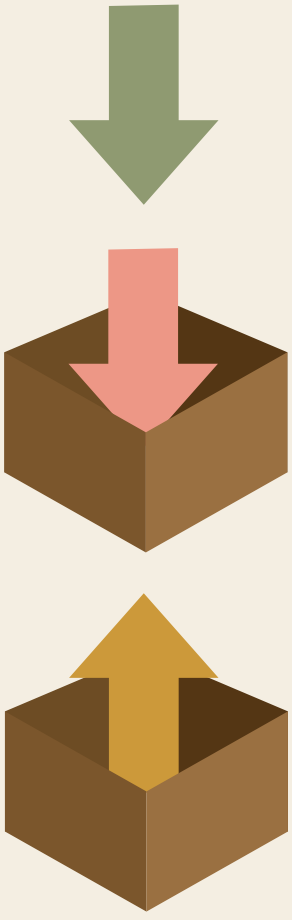
First Steps in Digital File Transfer and Storage

The transfer and storage of digital material is an increasingly common task for those managing and curating collections in universities, libraries, archives, museums, galleries and other institutions. Creating a workflow is a crucial first step to ensure that data is managed appropriately, is fully preserved during transfer, and continues to be managed and preserved as long as it is stored and used. Advance planning, selecting a suitable transfer method, using appropriate storage methods and carrying out follow-up preservation processes are all important parts of the workflow to consider. This guide introduces some of the important aspects of data management.



1. Preparing for data transfer

2. Transferring data

3. Storage, Use, Preservation

dri

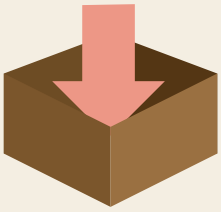
1.

Preparing for data transfer

- Liaise with your IT department. Communicate with technical experts to explain what you need to achieve and seek advice on best practice within your organisation.
- Preserve related documentation for storage in a separate folder with the collection or in your organisation's CMS (content management system) to capture the material's provenance (origin). For example, administrative history and contextual information related to a digital collection may form important accession information for the future. Note actions taken during the creation of the collection as well as any technical issues related to software or hardware. It is also essential that storage information be included so that a collection can be easily located.
- Identify sensitive data. Does the material contain any data governed by legislation or organisational privacy policies? For example, data containing personal or legal information may need extra security or access conditions attached.
- Will data be transferred from a different physical location or jurisdiction? This may influence your transfer choice or give rise to further legislative restrictions.
- Will data be transferred from an outside organisation that may be governed by legislation or other information management policies?

2.

Transferring data

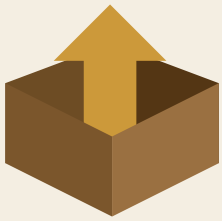


- Perform virus checks on all computers involved in the transfer process and ensure machines are protected with reliable and up to date security software.
- If possible, use write blockers to prevent alteration of file metadata and content.
- Select the most appropriate copying method: copy data individually, in batches or create a disk image. Copying data without creating a disk image is the simplest approach and may be adequate for transferring non archival data collections. Creating a forensic disk image makes an exact bit stream copy of a disk's contents, including forensic traces of use and content, while creating a logical image copies the contents and structure of the disk. Forensic disk imaging provides important information about file creation, transfer and editing enabling access to important information related to provenance.

- Make a copy of the file directory to document the structure of the folders and files you are copying and other basic information such as file names and sizes. This can be done in the command line or programmes can be downloaded for this purpose.
- Monitoring file fixity is crucial in order to detect whether files have been corrupted during transfer. File fixity can be verified using a file's checksum. A checksum is a unique sequence of letters and numbers which represents the exact bit stream of each individual file. Using specific software, the checksum of a file that has been transferred from one storage location to another can be compared to the checksum of the original file. If they do not match this indicates an alteration or corruption of the file. Use of file fixity tools during transfer is essential to retain the whole, authentic record.

3.

Storage, Use, Preservation



- Ensure that related documentation is updated, retained and stored appropriately.
- Following data transfer, file fixity should be scheduled to run at regular intervals and automatic checks can be run on batches of files. Regular checks will help to ensure that the integrity of the collection is maintained.
- Identify and validate the file formats. If selected for long term digital preservation, it is especially important to consider the suitability of the file formats in question. See DRI's Factsheet on **File Formats**[1].
- Appropriate user access rights and permissions should be set.
- Create two backup copies of the collection and separate from any working copies to ensure unintentional alteration does not take place.
- Backup copies should be stored separately in different locations on a secure network rather than on removable media. Backup methods should conform with your organisation's IT disaster recovery plan and, if necessary, trusted, robust forms of storage media should be selected.
- If a collection has been selected for long term digital preservation, it should be transferred to a trusted digital repository - remember laptops and hard drives are **not** archives!
- If a collection is deaccessioned, ensure that a secure, permanent method of deletion is used. Particular care should be taken when working with sensitive data.

[1] DRI File Formats Factsheet:
http://dri.ie/sites/default/files/files/Fact%20Sheet%20No%203%20File%20Formats%20v4.pdf
(accessed 10 August 2015)

Resources

The BitCurator project offers a suite of open source digital forensics and data analysis tools to help collecting institutions process born-digital materials, including forensic disk imaging tools.[2]

The National Archives of Australia developed the Digital Preservation Software Platform which offers free and open source software for digital preservation, including tools for file conversion and a checksum checker.[3]

AVPreserve is a consulting and software development firm. Its website features useful papers and tools, including Fixity, a utility for the documentation and regular review of stored files.[4]

The Library of Congress digital preservation website links to relevant tools and publications.[5] The Signal is the Library of Congress digital preservation blog which regularly shares relevant new developments.[6]

[2] http://wiki.bitcurator.net/, last accessed 14 October 2015
[3] http://dpsp.sourceforge.net/index.php, last accessed 14 October 2015
[4] https://www.avpreserve.com/, last accessed 14 October 2015
[5] http://www.digitalpreservation.gov/, last accessed 14 October 2015
[6] http://blogs.loc.gov/digitalpreservation/, last accessed 14 October 2015